WOODHOUSE
EXHIBIT 6

EXHIBIT E

Message

From: Ahuva Goldstand [[redacted]@meta.com]

Sent: 3/31/2023 11:55:25 PM

To: Rodrick Shepard [[redacted]@meta.com]; Jon Shepard [redacted]@meta.com]; Alex Yu [redacted]@meta.com]; George Orlin [redacted]@meta.com]; Angela Fan [redacted]meta.com]; Olga Rodstein [redacted]@meta.com]; Moya Chen [redacted]@meta.com]; Ahuva Goldstand [redacted]@meta.com]; Satish Mummareddy [redacted]@meta.com]; Melanie Kambadur [redacted]@meta.com]; Alex Boesenberg [redacted]@meta.com]; Sergey Edunov [redacted]@meta.com]; Robert Stojnic [redacted]@meta.com]; Elisa Garcia Anzano [redacted]@meta.com]; Sy Choudhury [redacted]@meta.com]; Lauren Cohen (Legal) [redacted]@meta.com]; Beau James [redacted]@meta.com]; Amanda Kallet [redacted]@meta.com]

Subject: Message summary [{"otherUserFbId":null,"threadFbId":5902043749833045}]

```
Alex Yu (3/31/2023 07:33:15 PDT):
>Question to the engineering team, when we release a chatbot that is not just English ( say German) , do
we have a big chatbot that is multi-language or we will have language specific chatbot?  For the datasets
do we need to do acquire multi-language ones?  For instance, for books, do we need to get all the
different language versions?  How do we think about this without the dataset acquisition work becoming
much bigger?   Thanks!

Melanie Kambadur (3/31/2023 07:37:35 PDT):
>we will likely have a hybrid: a big chatbot that is generically trained to understand multiple
languages, and then some different versions of it that are tuned to understand specific languages better.
I haven't dug in as much on what multilingual data we already have available, thought I know it is a
decent amount from various open data sources. Let me get back to you on a strategy here.

Sy Choudhury (3/31/2023 13:43:43 PDT):
>@Eng Team -
>
>
>Based om the sampling of the [redacted] books, *if* we were able to do a deal for 50K of their 300K
catalog, how close does that get us? Ie, is that just minimal-viable-product and we would still likely
need another 50K books for a Wave 2? Or our guess is that it gets us __some%__ to competitive?
>
>Thoughts? This is an important factor on how hard we may need to press on [redacted] management etc

Sy Choudhury (3/31/2023 13:44:33 PDT):
>To be more clear:
>
>> how close does that get us
>
>how close does that get us to a competitive LLM in the area of facts etc

Moya Chen (3/31/2023 14:05:12 PDT):
>50k seems a little small to me. Out of curiosity, is the thought here that [redacted] might be more
amenable to negotiating on a smaller volume of books or something else?

Moya Chen (3/31/2023 14:05:13 PDT):
>(Fwiw, it's not just the factuality from the books but also because we've got some work coming down the
pipeline for long context – ie, ways to let models read in large amount of text at once. We don't have
any super great, repetitive datasets for long context right now, and books would be incredibly helpful
for that.)

Sy Choudhury (3/31/2023 14:26:38 PDT):
>OK makes sense; the large amount of text per book etc

Melanie Kambadur (3/31/2023 14:35:01 PDT):
>I think it's also important which 50K. Is it still a diverse set of topics (better if it is)? Also some
of their 'books' are basically like graduate theses and shorter, while some are longer/higher quality
like full university textbooks. We'd prefer more of the latter.

Alex Boesenberg (3/31/2023 16:00:55 PDT):
>Question for Moya - for textbook and fiction book deliveries.  Do we ever want to use an API?  Or would
we always them to copy the XML files into AWS S3? Or some other internal storage we have?

Melanie Kambadur (3/31/2023 16:03:20 PDT):
>In case Moya is traveling now: we want whatever the vendor can get to us sooner, since we are flexible.
If I recall correctly, springer delivery was faster with an api so if that's true it's fine

Alex Boesenberg (3/31/2023 16:05:38 PDT):
>Ok
```

Alex Yu (3/31/2023 16:14:54 PDT):
>@Angela Fan, @Moya Chen @Melanie Kambadur, other than libgen what other datasets that you think will really help you but could not use today? Will be great to see a list like that. Thanks!

Melanie Kambadur (3/31/2023 16:16:46 PDT):
>Do you mean publicly available datasets that might be legally risky? Or do you mean reprioritizing the licensable datasets again in some way?

Alex Yu (3/31/2023 16:17:15 PDT):
>the first one.

Melanie Kambadur (3/31/2023 16:17:50 PDT):

Redacted - Privilege

Alex Yu (3/31/2023 16:18:29 PDT):
>how about books?

Melanie Kambadur (3/31/2023 16:19:09 PDT):

Redacted

Alex Yu (3/31/2023 16:32:11 PDT):
>Have you looked at Movie scripts? Any public site that is good?

Melanie Kambadur (3/31/2023 16:55:10 PDT):
>Ah yes simply scripts and internet movie script database

Melanie Kambadur (3/31/2023 16:55:25 PDT):
>Both look interesting